S1 Revision Notes

Key Words: Continuous, discrete, mean, median, mode, cumulative frequency tables, class width/midpoint, coding

- Mean: \( \bar{x} = \frac{\sum x}{n} \) or \( \frac{\sum f x}{\sum f} \)
- For large data: use coding to work out mean of coded data, then equate to coding for original mean
- Linear interpolation (for some x): using your bivariate data, sketch x against f(x). Plot some a where a < x and some b where b > x and consider the gradient of \( \frac{f(x-b)(a-x)}{b-a} \)
  - Used to calculate values in cumulative frequency tables
  - Combined mean of set A (size: n, mean: x) and of set B (size: m, mean: y) is given by \( \frac{n x + m y}{n + m} \)

Key Words: range, lower/upper quartiles, interquartile range, percentile

- Variance \( \sigma^2 = \frac{\sum x^2}{n} - \mu^2 = \frac{1}{\sum} \left( \frac{x}{\sum} \right)^2 \)
- Standard deviation: how much the members of a group differ from the mean value for the group
- For large data: use coding as a standard deviation of coded data
- Additive property: DOESNT affect the SD, but subtractive does! (Same to the SD!)
- Percentile of a variable is a value up to which xth Percentile \( \frac{100}{n} \) of the data is found (used to calculate percentile range)

Key Words: stem and leaf diagram, outliers, box plot, histograms, positive/negative skew,

- An outlier (typically) is any value:
  - Greater than the upper quartile + 1.5 x interquartile range
  - Less than the lower quartile – 1.5 x interquartile range
- Box plot: outlier, (lowest value), (lower quartile), (median), (upper quartile), (highest value), (outlier)
- Histograms (area is proportional to frequency):
  - Frequency density \( \times \) class width = frequency
  - Frequency density always on the y-axis
- Skew is calculated as follows
  - \( Q_2 - Q_1 < Q_3 - Q_2 = +ve \)
  - \( Q_2 - Q_1 > Q_3 - Q_2 = -ve \)
  - Mode < mean = +ve
  - Mode > mean = -ve
- Using \( x = \frac{3(\text{median-mean})}{\text{standard deviation}} \) if
  - \( x = 1 \) = perfect positive skew
  - \( x = -1 \) = perfect negative skew

Key Words: Venn diagrams, Mutually Exclusive, Exhaustive, Independent, Complementary, Conditional Probability

- Mutually Exclusive: events that cannot happen at the same time. e.g. passing and failing an exam
  - \( P(A \text{ or } B) = P(A) + P(B) - P(A \cap B) \text{ if NOT ME} \)
  - \( P(A \text{ or } B) = P(A) + P(B) \text{ if ME} \) \( P(A \cap B) = 0 \)
- Exhaustive: events where all possible outcomes are included. e.g. throwing a head or a tail on a fair coin
  - \( P(A \text{ or } B) = 1 \)
- Independent: one event has no effect on another event occurring e.g. throwing a 1 or a 2 on a fair dice

- \( P(A \text{ and } B) = P(A \cap B) = P(A) \times P(B) \text{ if independent events} \)
- Complementary: \( P(A') = 1 - P(A) \)
- Conditional Probability: \( P(A \text{ given } B) = \frac{P(A \cap B)}{P(B)} \)
- Multiplication rule: \( P(A \text{ and } B) = P(A) \times P(B|A) = P(B) \times P(A|B) \)
- Addition rule: \( P(A \text{ or } B) = P(A) + P(B) - P(A \cap B) \)

Key Words: independent (explanatory) variable, dependent (response) variable, residuals, regression line, interpolation, extrapolation

- A distance, e, from a point (on scatter diagram) to the best fit is called a residual
  - Outliers have relatively large residuals
  - The line of best fit aims to minimise \( \sum e^2 \) called the regression line of y on x
  - Equation of the regression line: \( y = a + bx \) where \( b = \frac{S_{xy}}{S_{xx}} \) and \( a = \bar{y} - b\bar{x} \)
- Interpolation: estimate values within the range of data
- Extrapolation: estimate values outside the range of data

Key Words: random variable, discrete random variable, cumulative frequency distribution

- For discrete random variable: \( \sum P(X = x) = 1 \)
- In a cumulative frequency distribution \( F(x) = P(X \leq x) \)
- \( E(X) = \sum P(x) \times X \)
- \( Var(X) = E(X^2) - E(X)^2 \)
- \( E(aX + b) = aE(X) + b \)
- \( Var(aX + b) = a^2Var(X) \)
- For a discrete uniform distribution over the values \( 1, 2, 3, \ldots, n \):
  - \( E(X) = \frac{n+1}{2} \)
  - \( Var(aX + b) = \frac{(n+1)(n+1)}{12} \)

Key Words: random variable, mean, standard deviation

- If \( X \sim N(\mu, \sigma^2) \)
  - \( Z = \frac{X - \mu}{\sigma} \) (called standardising)
  - \( \Phi(z) = P(Z < z) \)
  - \( P(\mu - Z \leq Y \leq \mu + Z) = \Phi(Y) - \Phi(X) \)